

ABSTRACT

An object of the present invention is to enable optimal clustering for many types of noise data and to improve the accuracy of estimation of a speech model sequence of input speech. Noise is added to speech in accordance with noise-to-signal ratio conditions to generate noise-added speech (step S1), the mean value of speech cepstral is subtracted from the generated, noise-added speech (step S2), a Gaussian distribution model of each piece of noise-added speech is created (step S3), the likelihoods of the pieces of noise-added speech are calculated to generate a likelihood matrix (step S4) to obtain a clustering result. An optimum model is selected (step S7) and linear transformation is performed to provide a maximized likelihood (step S8). Because noise-added speech is consistently used both in clustering and model learning, clustering for many types of noise data and an accurate estimation of a speech model sequence can be achieved.